Assembling
The Tree of Life (AToL)

Big business
In parsimony

Joseph K.-K. Li, Ph.D.

College of Science, Department of Biology

David J. Paper, Ph.D.

College of Business, Department of Business Information Systems

Kenneth B. Tingey, MBA, MPIA

College of Business, Department of Business Information Systems

Systems biology

- Systems biology
  - **Predictive, preventative & personalized** biology
  - Better care, lower cost, more medical/biological services
- Widely anticipated, as outlined by Lee Hood
  - Technology-led/genomics-based
  - We are modeling our SB approach after the NSF AToL Project -- **development of complete phylo*genetic* tree**
- We are using Utah-originated logic model to enhance efforts *to develop phylogeny-based taxonomies* with maximum **parsimony** and **functionality**

What can be done with phylogenetic trees?

- Organism **generation** and **growth** can be simulated and studied

- **Interactions** among organisms can be simulated, studied, and predicted

- Data on **individual organisms** can be stored for ready use
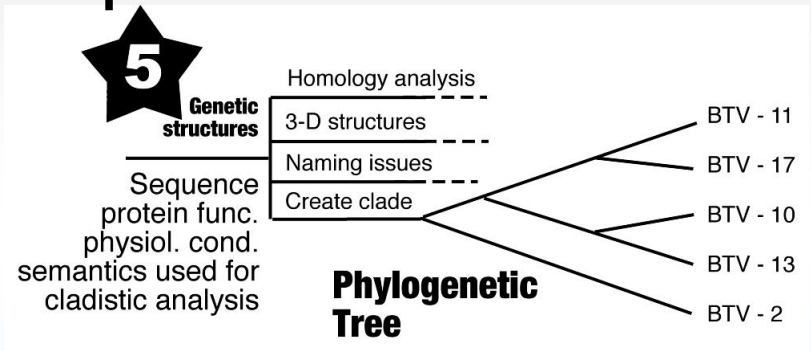
Systems biology

Systems Biology Companies		
Company	Rev.$Mil VC/Fed $Mil.	Product Service Mix
Beyond Genomics	- 21/12	B, C
BioSeek	- 8.4	B, C
Entelos	5.4 39.6	B
Genedata	- Yes	A
GeneGo	Yes Yes	B
Gene Network Sciences	Yes Yes/6+	A, C
Genomatica	Yes Yes/2.2	B
Ingenuity	Yes Yes/-	B
LineaGen [GenData]	- -	B, C
Predix/ Physiome	12.5+ 71/-	C

Sources: *Drug & Market Development Publishing*. 2004. Systems Biology: Key to unlocking the value within the omics revolution; company references.

- **Modeling** and **simulation** of biological systems
  – Science and pharma applications
  – Ongoing for at least a decade
- $1 billion by 2010/$100 million now, including DNA-based research
- Current competitors use **multi-component business models**
  A. Software and/or data generation
  B. Services based on the same
  C. Drug targets and diagnostic markers

Parsimony

- Systems biology is dependent on <u>elimination of any and all irrelevant complexity</u> -- ***maximum parsimony along with functionality are the keys***

- **Only biologists** can make these critical distinctions

Problems with traditional IT approaches to biology

- **Many arbitrary models**
  - Frege's predicate calculus
  - Complex programming languages
  - Runaway database table creation

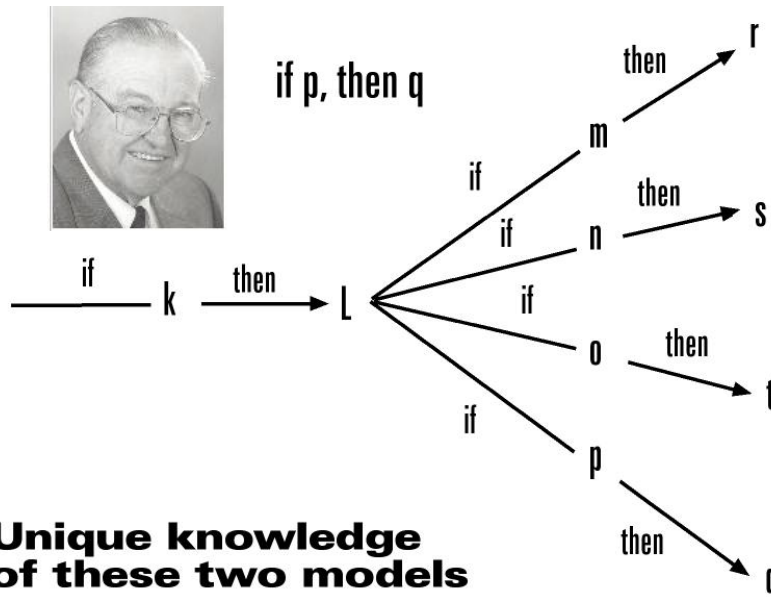- Prevalent IT models are *difficult to learn*, with *high maintenance costs*

- Parsimony suffers

Parsimony at fundamental levels

Twin factors underlying the Li, Paper, Tingey approach to Assembling the Tree of Life (AToL)

Generative taxonomies (**AGT**)

Basis & structure for all logical relationships

if p, then q

if k then L

if → then → m → r
if → then → n → s
if → then → o → t
if → then → p → q

Unique knowledge of these two models

Phylogenetic system and model design by scientists for scientists and by industrialists for industrialists

Initial products by 2007

DNA/RNA double helix

Uracil (U) ⟷ (A)

Basis & structure of all living things

cytosine (C) to guanine (G)
adenine (A) to thymidine (T)

C — G
G — C
A — T
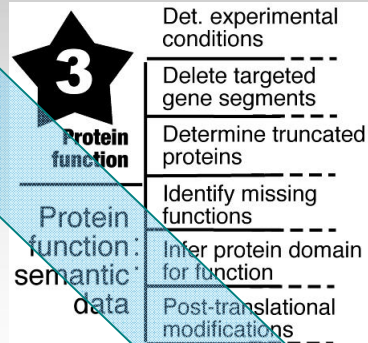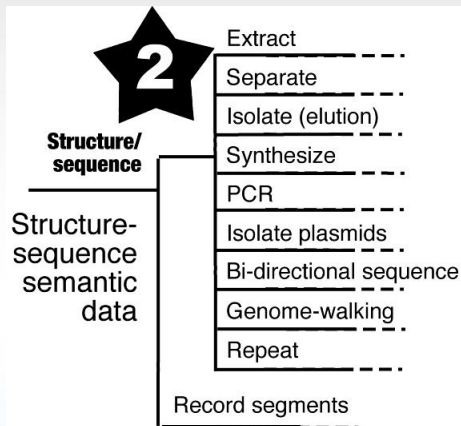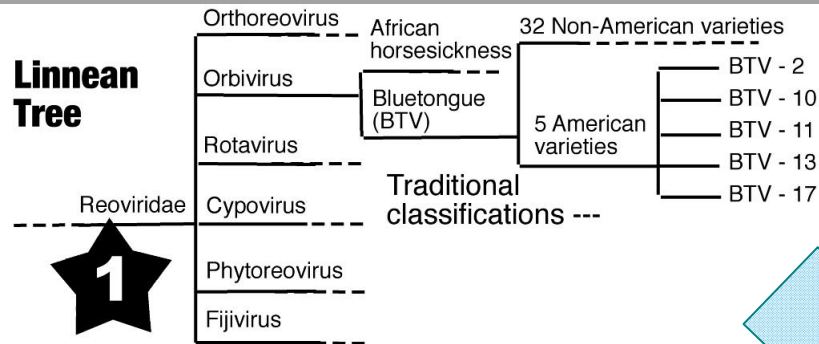A — T
G — C
G
A
A
T
G
G
C — G
T — A
C
G — C
A — T
A — T
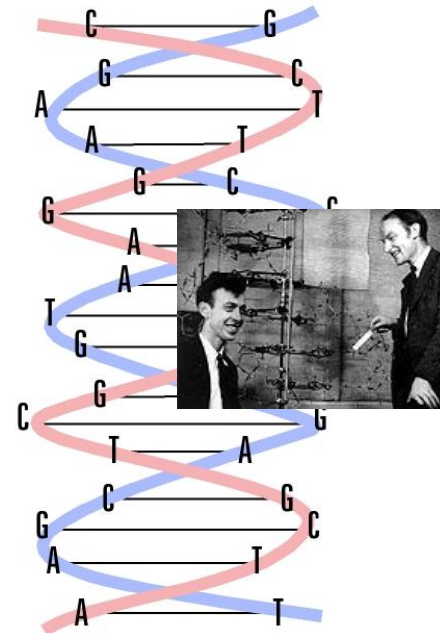
AGT a new paradigm

- Link **all available** data from *analytical instruments* and *assay output* into a single, secure, parsimonious system

- *Allen Generative Taxonomy* (**AGT**) model very **user/designer-friendly** (also existing Tingey "Big E, little e" patent application)

- Will have substantial <u>gains in productivity and efficiency</u> along with <u>significant time and cost saving</u> -- *achieving parsimony at the core of both logic and genetic structure*
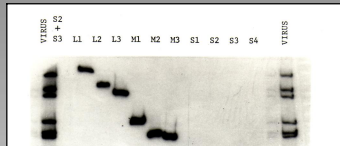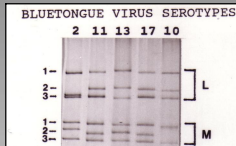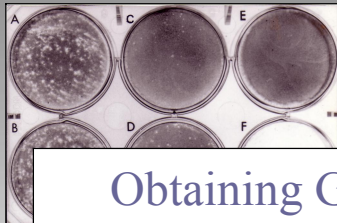
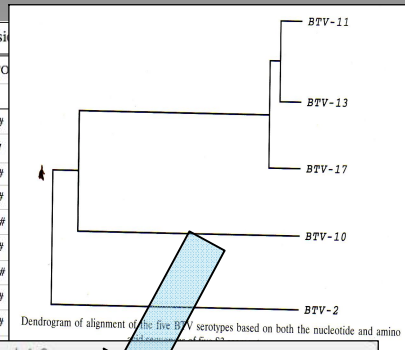# From descriptive to phylogenetic trees
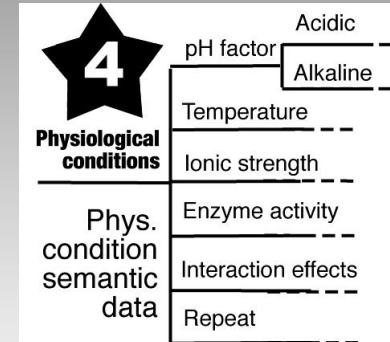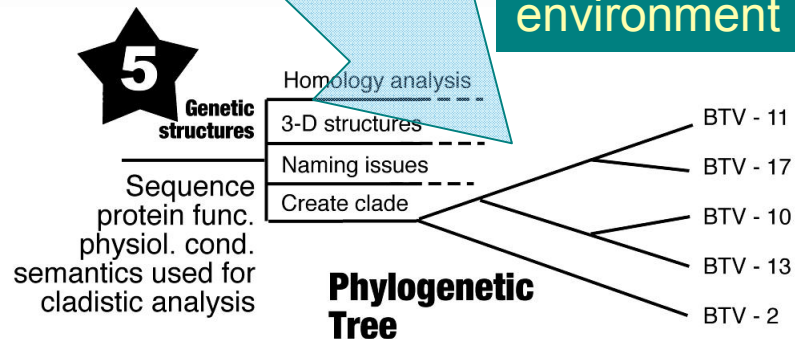
Under **AGT**, manage taxonomies and processes in same environment

Nature of the process

Obtaining Genomic Sequences from Bluetongue
Virus ds-RNA Gene Segments
Isolation from infected BHK-21 cells via density centrifugation

- **Extract** BTV viral ds-RNA genome seg...
- **Separate** ds-RNA genome segment by g...
- **Isolate** and **elution of** ds-RNA segments...
- **Synthesize** cDNA using terminal primer...
- **Insert** PCR products into plasmids that v...
- **Isolate** plasmids serving as template for...
- **Bi-directional sequence** with terminal p...
segments
- **Use "genome walking"** with new prime...
- **Repeat** 3-4X with different batches of v...
- **Record gene sequences** for **comparativ...
alignment/phylogenetic analysis**

From descriptive to phylogenetic trees

A discovery, not hypothesis-driven approach

From descriptive to phylogenetic trees

AGT is well-formed and robust – we are positioning it for science and business

Project plans

- **Three year plan** for _generative phylogenetic taxonomies_ of BTV, Avian Flu, and SARs viruses, then Anthrax, mice, rats, and sheep

- **Partnership with others** at USU/Utah, Thailand, Hong Kong, NEON, IBM

- **Supercomputer capacity** partnering with NSF, Dell, and IBM

- _Co-location of hardware_ with **National LambdaRail** in Ogden/SLC

- Commercial offerings by 2007
  - Pages 16-17 of COE proposal
  - Product code names: **SuperTree**, **Filo**, **GenoTree**, **ViruTrak**

Long-term goals

- Combine/record the *integrated data and process types* with **maximum parsimony** (sequences, expression profiles, and biochemical/biological pathways, etc.)

- Create a *unified exploration and practice environment* based on **AGT**

- Help **biologists** to **make rapid advances** in their research and practice

- Support ***more fluid enterprise models*** for **biology** and **other knowledge-intensive environments**